(2)

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

AD-A233 889 ORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS N/A
2a. SECURITY CLASSIFICATION AUTHORITY N/A	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution is unlimited
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE N/A	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) USAFSAM-JA-90-17	5. MONITORING ORGANIZATION REPORT NUMBER(S) N/A
6a. NAME OF PERFORMING ORGANIZATION USAF School of Aerospace Medicine 6b. OFFICE SYMBOL (If applicable) VNBA	7a. NAME OF MONITORING ORGANIZATION N/A
6c. ADDRESS (City, State, and ZIP Code) Human Systems Division (AFSC) Brooks AFB TX 78235-5301	7b. ADDRESS (City, State, and ZIP Code) N/A
8a. NAME OF FUNDING/SPONSORING ORGANIZATION Aerospace Medicine 8b. OFFICE SYMBOL (If applicable) USAFSAM/VNBA	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
8c. ADDRESS (City, State, and ZIP Code) Human Systems Division (AFSC) Brooks AFB TX 78235-5301	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO: 62202F; PROJECT NO.: 7930; TASK NO: 14; WORK UNIT ACCESSION NO.: 39

11. TITLE (Include Security Classification)
Errors in Measurement of +Gz Acceleration Tolerance

12. PERSONAL AUTHOR(S)
Ludwig, David A., and Krock, Larry P.

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Final	FROM 87-06-01 TO 90-02-01	90 February	20

16. SUPPLEMENTARY NOTATION
None

17. COSATI CODES		
FIELD	GROUP	SUB-GROUP
06	10	

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
Acceleration, Acceleration tolerance
Statistical analysis, Research design

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

Most acceleration studies estimate a subject's G-level tolerance by taking only one determination (test) for a given condition. The purpose of this study was to examine the error structure and reliability of an individual's acceleration tolerance and to provide design considerations for future experimentation. A hierarchical (nested) design was used to estimate the source of variation in measuring G-level tolerance. Six males rode relaxed in the USAF School of Aerospace Medicine human-use centrifuge and were exposed to a 0.1G/s onset rate profile until greyout. Each subject was tested on three randomly selected days with three repeated determinations within a day. This design allowed for an estimate of both day-to-day and measurement error within a testing session. A single +Gz tolerance determination was found to be moderately unreliable (rel coef. = 0.80). Under the best of circumstances a subject's G-level tolerance cannot be estimated with any more accuracy than about ±0.3 G with 95% confidence. This degree of accuracy can only be obtained with multiple measurements.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT	21. ABSTRACT SECURITY CLASSIFICATION
☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS	Unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL Larry P. Krock, Ph.D.	22b. TELEPHONE (Include Area Code) (512) 536-3521 22c. OFFICE SYMBOL USAFSAM/VNBA

DD Form 1473, JUN 86 *Previous editions are obsolete*

SECURITY CLASSIFICATION OF THIS PAGE
Unclassified

TECHNICAL NOTE

Errors in Measurement of $+G_z$ Acceleration Tolerance

DAVID A. LUDWIG, Ph.D., and LARRY P. KROCK, Ph.D.

LUDWIG DA, KROCK LP. *Errors in measurement of +Gz acceleration tolerance*. Aviat. Space Environ. Med. 1991; 62:261–5.

Most accleration studies estimate a subject's G-level tolerance by taking only one determination (test) for a given condition. The purpose of this study was to examine the error structure and reliability of an individual's acceleration tolerance and to provide design considerations for future experimentation. A hierarchical (nested) design was used to estimate the sources of variation in measuring G-level tolerance. Six males rode relaxed in the USAF School of Aerospace Medicine human-use centrifuge and were exposed to a 0.1 G/s onset rate profile until greyout. Each subject was tested on three randomly selected days with three repeated determinations within a day. This design allowed for an estimate of both day-to-day and measurement error within a testing session. A single $+G_z$ tolerance determination was found to be moderately unreliable (reliability coefficient = 0.74). Under the best of circumstances a subject's G-level tolerance cannot be estimated with any more accuracy than about ±0.3 G with 95% confidence. This degree of accuracy can only be obtained with multiple measurements.

AEROSPACE physiologists and engineers have endeavored for some time to counter the effects of $+G_z$ acceleration by suggesting maneuvers or designing devices that function to maintain adequate perfusion of the brain. To test the effectiveness of a proposed countermeasure, the maneuver or device is compared to an established standard. Results of such comparison often based on the $+G_z$ tolerance of human subjects. [Acceleration stress tolerance can be divided into two components: a level component, or the momentary attainment of a peak magnitude of stress (e.g., $+8.7\ G_z$), and duration component, or how long a specified level(s) of stress can be endured (i.e., time at G). Although the present study is concerned with the peak level component of tolerance, the findings may have application to the duration component as well.]

The recorded G-tolerance obtained from a single test is only an estimate of that individual's true G-tolerance. This number, in addition to the "true" G-tolerance, contains one or more sources of inconsistencies or errors. The purpose of a well conceived experimental design is to make certain that the recorded measurement of G-tolerance is as accurate as possible. Subject-to-subject, measurement, and within-subject (biological) variation are frequently encountered sources of error in acceleration research.

Between-subject response differences are well known. This source of variation is accounted for in most studies by using within-subject designs. [Within-subject designs refer to experiments where subjects are crossed with treatments (each subject receives every treatment) as opposed to between-subject designs where subjects are nested within treatments (each subject receives only one treatment).] Further, due in part to the physical demands of exposure to increased levels of acceleration and to the moderately strict medical requirements, human-use acceleration research is limited to small groups of healthy volunteers. The experimenter must therefore control error variance by design rather than by sample size.

Measurement error, inaccuracies in the technique of measuring a response, can add considerably to total variation. Acceleration researchers have tried to control this source of error by standardizing centrifuge performance characteristics and by improving upon the validity of methods for determining the tolerance endpoint (1,2,7,8,10). Subjective end points have been a source of controversy in acceleration research for some time, and will probably remain so in the future. Several methods to improve the reliability of subjective end point measurement have been proposed (5,10,11,17) and numerous objective G-tolerance measurement techniques have been offered as alternatives to subjective criteria (13,14,16).

Surprisingly, given the large investment in attempting

From the University of North Carolina, Department of Mathematics and Statistics, Greensboro, NC (D. A. Ludwig) and the USAF School of Aerospace Medicine, Crew Technology Division, Brooks AFB, TX (L. P. Krock).

This manuscript was received for review in July 1990. The revised manuscript was accepted for publication in August 1990.

Address reprint requests to: Larry P. Krock, Ph.D., Crew Technology Division, Acceleration Effects Function, USAF School of Aerospace Medicine (AFSC), Brooks AFB, TX 78235-5301.

to improve the measurement device, acceleration research has focused very little attention upon controlling within-subject variability. Only three studies have attempted to address this issue. Glaister and Hall (12) discussed the possibility of diurnal variation in human tolerance to sustained +G_z. Whinnery and Jackson (19) described the variance of rider response to a specific medical evaluation protocol and Stauffer (18) considered individual tolerance fluctuations over differing G-induced symptomatic states.

If researchers are concerned with improving the accuracy of determining an individual's acceleration tolerance, it is important for them to realize that various types of errors are components of every observed value. A better understanding of the contribution of these error components is only possible when they are simultaneously evaluated within a single sample.

The purpose of this study was to examine the error structure and reliability of an individual's acceleration tolerance and to provide design considerations for future experimentation.

METHODS AND MATERIALS

Subjects: Six males from the Brooks Air Force Base acceleration panel volunteered for this experiment. Ages ranged from 22 to 31 years with an average of 25.8 years. All subjects were in good health and were experienced centrifuge riders.

Centrifuge protocol: Subjects rode the centrifuge in an upright (13° backangle) seat without G-trousers and were instructed to remain as relaxed as possible. Electromyography of the quadriceps and the gastrocnemius muscles provided a method for the investigators to monitor muscle tension and remind the subjects to remain relaxed. Onset rate was set at 0.1 G/s.

Subjects rode the centrifuge until they experienced either 100% peripheral or 50% central visual field loss. Although subjects terminated each run based upon a subjective evaluation of their visual field loss, an objective record was obtained by way of a subject-controlled high-resolution visual field limit tracker (Fig. 1). This device consists of a curvilinear array of 120 white lights located at 1° increments around a 76-cm radius. Fixating on a central red light, subjects "balanced" the white lights on the edge of their visual field by triggering a micro-switch on the gondola joystick. Following each run, the objective record was used to verify the subjective endpoint. A more complete description of the apparatus as it was used in this study may be found in an article by Gillingham and McNaughton (11). A similar apparatus is described by Cammarota (5).

Each subject was tested in the morning on three different days within a 10-week period. A minimum interval of at least 1 d between tests was provided. On each of the 3 d, three measurements (tests) were taken. A 15-min rest interval was given between tests. A small pilot study indicated that this rest period was sufficient to eliminate order effects within the testing session. Heart rate was used to verify that the subject had returned to base line levels. Once the subject had terminated the test, the maximum +G_z value was recorded for use in the statistical analysis.

Statistics: A hierarchical (nested) design was employed to identify the sources of variation in the measurement of +G_z tolerance. A graphic illustration of what is being estimated is given in Fig. 2. The average +G_z tolerance of the given population is an unknown value (U) that would be realized if every subject within that population could be tested. The average difference of each subject's observed value (Y) from U would be used to calculate the error in estimating U (E_t). This total error is composed of three parts: subject-to-subject error (E_s), time-to-time (i.e., day-to-day) error (E_d) within a subject, and measurement error (E_m) within a given period of time for any particular subject. Although the accuracy of the estimation of U is important, more important to researchers conducting acceleration experiments is the estimation of a single subject's +G_z tolerance (S). Two sources of error, day-to-day and measurement, contribute to the variation of an individual tolerance. Similarly, tolerance for any given subject on any given day is only affected by error within the measurement process.

If the squared errors are summed and averaged they become variances. The hierarchical design allows for the estimation of these variances via the analysis of

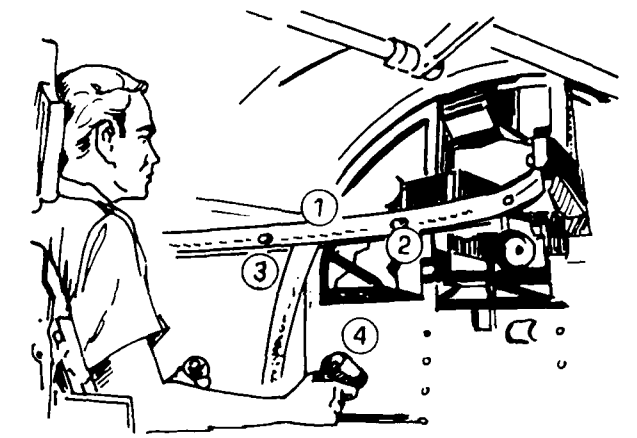

Fig. 1. Subject operating the USAFSAM high-resolution visual field limit tracker in the horizontal mode. 1) Light bar containing 121 equally spaced lamps in 120° arc. 2) Central red light for fixating gaze. 3) Centrally moving white lights (50°/sec). 4) Subject's direction-reversing response switch. (Permission to reproduce this illustration was obtained from Gillingham and McNaughton (11).)

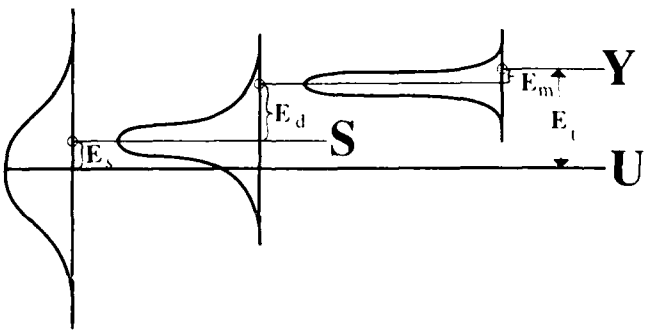

Fig. 2. Decomposition of the total error in estimating +G_z tolerance. Y = observed value, U = population mean +G_z tolerance, E_s = subject-to-subject error, E_d = day-to-day error within a subject, E_m = measured errors within a day and subject, $E_t = E_s + E_d + E_m = Y - U$. In terms of variance, $\sigma^2_o = \sigma^2_s + \sigma^2_d + \sigma^2_m$.

variance. Once these variance components are estimated they can be used to determine the degree of attention each source of error should be given in order to achieve a reliable estimate of $+G_z$ tolerance. A more detailed description of this statistical method is given by Box, Hunter, and Hunter (4).

RESULTS

The results of the analysis of variance and variance component estimation are presented in Table I. The large variance component associated with subject-to-subject differences is not surprising and is to be expected, given the wide diversity between humans. Differences between subjects accounted for 74% of the total observed variation in $+G_z$ tolerance. Day-to-day variation represented 16%, and measurement error accounted for the remaining 10%. Since reliability is defined as the ratio of true score variance to observed score variance, an estimate of the reliability of a single determination can be obtained by the ratio of the subject's variance component to the total variance component. The result is a reliability coefficient of 0.74, which is generally considered unacceptable by the scientific community. The results indicate that day-to-day variation and measurement error (test-to-test variation) should be considered when determining a subject's $+G_z$ tolerance. Variance components, in and of themselves, provide minimal information except in the form of relative magnitude. Their utility is in the estimation of precision. The standard error of a subject's $+G_z$ tolerance is given by the formula:

$$\sqrt{\frac{(\sigma m)^2}{(\#\text{ of d})(\#\text{ of tests per d})} + \frac{(\sigma d)^2}{(\#\text{ of d})}}$$

Using the variance component estimates from Table I in the above formula, the variation around a single determination is ±0.39 G (d = 1, tests = 1). An approximate 95% confidence interval (2 S.E.) would place the accuracy of this single test at ±0.78 G. The width of this interval is over 1.5 G, which is probably unacceptable for most scientific and clinical applications.

The standard error of a subject's $+G_z$ tolerance can be reduced by increasing the number of days on which a subject is evaluated and/or increasing the number of tests run within a particular day. The number of days and the number of tests can be substituted into the formula and an estimate of precision can be obtained around the mean of these multiple observations. For example, if a subject was measured on two separate days and two tests were run on each of these days, the standard error around the mean of these four observations would be ±0.25 G and the approximate width of the corresponding 95% confidence interval would be ±0.5 G. It could then be concluded with 95% confidence that a subject's "true" $+G_z$ tolerance value was within ±0.5 G of the estimated value.

Table II provides standard errors for several combinations of days and tests. Three aspects of this table are especially worth noting. First, the standard errors decrease at a faster rate when days are increased rather than tests, since more within-subject variation can be attributed to day-to-day variation than to measurement error. Second, it is not possible to reach an adequate degree of precision (±0.5 G, 95% confidence interval) when a subject is evaluated on a single day, regardless of the number of tests run within that day. And, third, the cost of added precision above ±0.3 G is not practical for most experimental situations.

These points are further illustrated by observing the change in the reliability of a single test when one accounts for the day-to-day variability or measurement error. When day-to-day variation is removed, the reliability of a single test increases to 0.88 as compared with a reliability coefficient of 0.82 when measurement error is removed. The reliabilities calculated above represent optimistic projections, since any source of error can never be completely eliminated.

TABLE I. HIERARCHICAL ANALYSIS OF VARIANCE AND VARIANCE COMPONENT ESTIMATIONS.

Source of variation	df	MS	σ^2
Subject	5	4.496	0.46
Days/Subject	12	0.359	0.10
Tests/Days/Subject	36	0.056	0.06
Total (corrected)	53		0.62

Note: "/" represents "nesting"; i.e., days within subjects.

DISCUSSION

Given its high standard error and low reliability, researchers should be cautioned against using only one determination of a subject's $+G_z$ tolerance. When a

TABLE II. STANDARD ERRORS FOR DIFFERENT COMBINATIONS OF DAYS AND TESTS.*

		DAYS					
		1	2	3	4	5	6
TESTS	1	.39	.28	.23	.22	.18	.16
	2	.36	.25	.21	.18	.16	.15
	3	.34	.24	.20	.17	.15	.14
	4	.34	.24	.19	.17	.15	.14
	5	.33	.23	.19	.17	.15	.14
	6	.33	.23	.18	.16	.15	.13

Estimation of at least ±0.3 G with 95% confidence

Estimation of at least ±0.3 G with 95% confidence

* Refers to tests within days.

single measurement is used, an estimated 26% of what is observed is not associated with the "true" +G_z tolerance of the individual. Control of this error variance which arises from biological variation and measurement error is of primary concern when a clinical evaluation of a subject's G-tolerance is needed. This added variance may also hinder attempts at experimental comparisons when G tolerance is used as the dependent variable. In addition, high subject-to-subject variability and the small number of subjects historically associated with acceleration research necessitate the use of within-subject designs which must rely primarily on multiple determinations to increase their efficiency.

Researchers should not be overly concerned with reducing measurement error as a method of increasing the reliability and precision of a single G tolerance value, since 63% of the error in estimating a subject's +G_z tolerance can be attributed to day-to-day variation. This is not to say that measurement error should be ignored, but suggests that the control of day-to-day variation would be a more efficient approach. Under the very best of circumstances, a subject's G-tolerance cannot be estimated with any more accuracy than about ±0.3 G with 95% confidence. This degree of accuracy can only be obtained with multiple measurements. How and when these measurements are taken must be based on cost, time, and the experimental situation. The results of this study suggest that when all else is equal, additional measurement over several days is preferable to additional tests within a day.

It is certainly possible that the day-to-day component can be further partitioned into time of day variation (a.m. vs. p.m.). Further research will be needed and should be undertaken to address this issue. Studies concerned with the causes of daily variation in G tolerance also seem warranted since many of the suggested factors (illness, fatigue, etc.) affecting day-to-day G-tolerances are described only anecdotally (3,9,18). Until more information is available on day-to-day changes in G tolerance, researchers will be well advised to consider day-to-day subject variation as an important source of error.

Although measurement error was the smallest source of variation in this study, reducing it should not be overlooked as a means of increasing accuracy. Measurement error was probably underestimated in the present study because all subjects were trained and familiar with the physiological sensations associated with the subjective end point. If untrained subjects had been used, the measurement error component would have probably been larger. Therefore, when naive subjects are used, objective endpoints would be preferred over subjective measurement. The results suggest that, in well-trained subjects, subjective endpoint techniques can be regarded as valid experimental procedures.

Increasing sample size during experimental evaluation is by far the best way of increasing the efficiency of an experimental investigation. However, if the number of subjects is fixed and relatively small, efficiency can only be improved by reducing biological variation and measurement error. Examination of a subset of the data from a recent article by Krutz, *et al.* (15), helps to illustrate this point. A comparison of two types of G trousers yielded a difference of 0.4 G when five subjects were tested in a within-subjects design (relaxed, G valve active). If a statistical hypothesis test had been performed to test for population differences (alpha = 0.05), a non-significant statistical difference would have been found ($t = 2.51$, $df = 4$, $p = 0.066$). This would have then indicated that the sample size was not sufficient to place an adequate degree of confidence in the observed 0.4-G difference. This probably would not have been the case if subjects were tested twice for each G-trouser condition and the subjects' mean values across these two tests used in the statistical analysis. This is not to say that the differences observed by Krutz, *et al.*, are not real, only that if statistical significance tests had been used, a sample size of five would have been inadequate since only one determination of G-tolerance was taken. Although multiple determinations per subject per treatment condition require more testing, the additional effort is certainly more favorable than wrongly concluding that a particular treatment was ineffective.

Any statistical estimate is subject to error. The variance components determined from this study are no exception. The small sample size of six subjects is certainly a limitation. As with any study, the present results are specific to the methodologies employed. The rate of onset, seat configuration, and subject state (relaxed) are all aspects that could possibly change the error structure of an observed +G_z tolerance. However, studies using much larger samples, different types of subjects, and different onset rates do provide estimates of subject-to-subject variability and total within-subject variability that are comparable with the present investigation (6,19). This provides, at least circumstantially, evidence that the results of this study may be extended to apply to a number of different situations.

This investigation was undertaken not only to gain additional information on how humans respond to +G_z acceleration, but also to provide information that researchers can use to design more efficient acceleration experiments. The results are presented as guidelines; subsequent implications must be considered in light of the experimental situation. Given the small number of acceleration laboratories, the small number of subjects willing to volunteer for such experiments, and the decreasing benefits of additional anti-G devices and maneuvers as human subjects reach the limits of their physiological capabilities, investigators must be increasingly aware that the odds of uncovering true experimental differences are stacked against them. Given this situation, the time it takes to refute published findings resulting from inadequate designs is far greater than for most other areas of aerospace research.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the support and helpful discussions of Dr. Russell R. Burton. Funding for this research was provided in part by the Air Force Office of Scientific Research and the USAF School of Aerospace Medicine. The voluntary informed consent of the subjects used in this research was obtained in accordance with AFR-169-3. All human experimentation was approved by the USAF School of Aerospace Medicine Advisory Committee on Human Experimentation prior to study initiation.

REFERENCES

1. Anonymous. Standardized centrifuge G-Stress profiles for medi-

cal evaluation of aircrew members. Washington, DC: Air Standardization Coordinating Committee, monitored by International Standardization Office, Headquarters USAF, 1982; Advisory Publication (ADV PUB) 61/26A.
2. Anonymous. Minimum requirements for selection, training and employment of aircrew in high sustained "G" environment. Brussels, Belgium: Military Agency for Standardization, NATO, 1981; Standardization Agreement (STANAG) No. 3827.
3. Balldin UI. Factors influencing G-tolerance. Clin. Physiol. 1986; 6:209–19.
4. Box GE, Hunter WG, Hunter JS. Statistics for experimenters. New York: John Wiley and Sons, 1978.
5. Cammarota JP. A training method for the dynamic tracking of peripheral vision limits. Warminster, PA: Naval Air Development Center, 1985; Report No. NADC-86072-60.
6. Cochran LB, Gard PW, Norsworthy ME. Variations in human G tolerance to positive acceleration. Pensacola, FL: U.S. Naval School of Aviation Medicine, 1954; Project No. NM 001059.02.10.
7. Coburn KR. Physiological endpoints in acceleration research. Aerosp. Med. 1970; 41:5–11.
8. Gillingham KK. G tolerance standards for aircrew. In: Preprints of the Aerospace Medical Association 1978 Annual Scientific Meeting. Washington, DC: Aerospace Medical Assoc. 1978:77–8.
9. Gillingnam KK. High-G stress and orientation stress: physiologic effects of aerial maneuvering. Aviat. Space Environ. Med. 1988; 59(suppl):A10–20.
10. Gillingham KK. G-tolerance standards for aircrew training and selection. Brooks AFB, TX: USAF School of Aerospace Medicine, 1986; Technical Report No. 86-12.
11. Gillingham KK, McNaughton GB. Visual field contraction during G stress at 13°, 45° and 65° seatback angles. Aviat. Space Environ. Med. 1977; 48:91–6.
12. Glaister DH, Hall GM. Circadian variations in tolerance to +Gz acceleration. Physiologist 1982; 25(suppl):S25–8.
13. Jaron D, Moore T, Shankara-Reddy BR, Hrebien L, Kepics F. Reflectance photoplethysmography as an adjunct to assessment of gravitational acceleration tolerance: preliminary findings. Aviat. Space Environ. Med. 1987; 58:604–12.
14. Knutson JR, Wood EH. Use of an improved photoelectric plethysmograph for determination of blood loss from the ear on the human centrifuge. Am. J. Physiol. 1948; 155:448–9.
15. Krutz RW, Burton RR, Forster EM. Physiological correlates of protection afforded by anti-G suits. Aviat. Space Environ. Med. 1990; 61:106–11.
16. Krutz RW, Rositano SA, Mancini RE. Comparison of techniques for measuring +Gz tolerance in man. J. Appl. Physiol. 1975; 38:1143–5.
17. Rogge JD. Relation of signal light intensity to physiologic end points during +Gz acceleration. Brooks AFB, TX: USAF School of Aerospace Medicine, 1986; Technical Report No. 68-38.
18. Stauffer FR. Acceleration problems of naval air training: I. Normal variations in tolerance to positive radial acceleration. Pensacola, FL: U.S. Naval School of Aviation Medicine, 1952; Project No. NM 001059.02.09.
19. Whinnery JE, Jackson WG. Reproducibility of +Gz tolerance testing. Aviat. Space Environ. Med. 1979; 50:825–8.